ESPCI PARIS
EDUCATION SCIENCE INNOVATION

Paris, the 29 March 2025

The dissertation *"Ensemble encoding of self- and other-related behaviors in the medial prefrontal cortex"* by Konrad Danielewski, conducted at the Nencki Institute of Experimental Biology under the supervision of Prof. Ewelina Knapska and Dr. Kacper Kondrakiewicz, presents a rigorous investigation into how the medial prefrontal cortex (mPFC) encodes both self- and other-related behaviors in rats. The study is highly relevant to current neuroscience, particularly in the domains of social cognition and neural population coding, as it addresses whether observed behaviors are represented at the level of individual neurons, as posited by the classical mirror neuron hypothesis, or at the level of distributed population activity. The work is methodologically sound, integrating advanced behavioral paradigms, high-density electrophysiological recordings, and computational analyses to examine the neural mechanisms underlying social behavior.

The introduction of the dissertation is particularly well-written, providing a comprehensive review of the literature on rodent social behavior, the functional role of the mPFC, and the conceptual tension between single-neuron and population-level coding. The candidate demonstrates a strong command of the relevant theoretical frameworks, integrating perspectives from neuroscience, ethology, and computational modeling. I found the introduction especially valuable in that it presents a coherent and structured vision of the field—one that, while open to debate, is conceptually consistent and intellectually compelling.

Methodologically, the study makes use of a broad array of techniques, including Neuropixels electrophysiological recordings, pose estimation using DeepLabCut, dimensionality reduction approaches and decoding analysis. The behavioral paradigm, adapted from social buffering protocols, is well-designed, allowing the author to investigate how rats integrate information about a partner's behavior to adjust their own responses. The use of machine learning classifiers to decode neural activity adds a quantitative dimension to the analysis, enabling an objective evaluation of how population activity relates to distinct behavioral states.

The results provide convincing evidence that the mPFC encodes both self-generated and observed behaviors at the population level rather than through individual neurons acting as mirror neurons. The analyses show that observed behaviors are associated with distinct neural trajectories, supporting the view of a flexible, context-dependent representational scheme. Furthermore, the study demonstrates that behavioral states such as freezing and rearing are encoded independently of simple motor kinematics, suggesting that the mPFC supports higher-order representations rather than reflecting low-level motor output. The decoding analyses indicate that both self- and partner-related behaviors can be classified above chance, highlighting the predictive structure embedded in population activity.

The decoding results are clear and robust. Classification accuracy typically reaches around 60% for a six-class problem, far above the chance level of 16.6%, and reflects structured, behaviorally relevant information in the recorded neural activity. This is particularly notable given the complexity and ecological validity of the behavioral paradigm. However, the decoding performance remains moderate, and some behavioral classes are more reliably identified than others. This raises important questions about the specificity and distribution of encoding across the mPFC population. While the analyses clearly demonstrate the presence of social and behavioral information in the signal, they do not yet define a mechanistically transparent model of how such

ESPCI Paris
Laboratoire Plasticité du cerveau - UMR 8249
10 rue Vauquelin
75231 Paris Cedex 05
Tel : +33 1 40 79 47 56
Fax : +33 1 40 79 47 57
www.espci.fr

representations are formed and used. The findings thus point to a complex and distributed neural code that warrants further investigation.

The discussion situates these results effectively within the broader literature on social cognition, population dynamics, and decision-making. The author persuasively argues that population-level encoding provides a more flexible and robust mechanism than single-neuron mirroring. A particular strength of this section is the theoretical reflection on how the mPFC integrates internal and external information in a socially relevant context. The findings are framed within broader constructs such as social learning, associative processing, and geometrical representations of neural activity, contributing to a compelling interpretive framework. Nonetheless, certain points might have been explored in greater depth—for instance, the potential influence of other brain regions (e.g., the amygdala or hippocampus), the impact of motivational and attentional factors, or the possible role of temporal coding schemes.

One broader conceptual question that emerges from this work is the precise functional role of the mPFC in social cognition. While the dissertation convincingly demonstrates that this region encodes both self- and other-related behaviors, it remains unclear whether such encoding reflects passive cue extraction, active behavioral prediction, or internal model construction. Does the mPFC merely represent observed actions, or does it contribute to building interpretable models of others' behavior and intentions? This distinction has important implications for understanding the computational functions of prefrontal circuits in social contexts. Nevertheless, the discussion remains one of the strongest aspects of the dissertation and places the findings within a well-developed and innovative scientific narrative.

Overall, the dissertation makes a clear and original contribution to the field of social neuroscience and neural population coding. It challenges existing theories on mirror neurons and supports a more dynamic, population-based framework for understanding how social behaviors are encoded in the brain. The work demonstrates the candidate's capacity to conduct independent, methodologically sophisticated research and to engage critically with complex theoretical issues. The findings are novel, well supported, and likely to stimulate further research on neural population dynamics and social cognition. In view of the scientific merit, originality, and methodological rigor of the work, I strongly recommend that Konrad Danielewski be admitted to the subsequent stages of the doctoral defense. Moreover, I consider this dissertation to be of outstanding quality and support its distinction.

Philippe Faure
Research Director CNRS
Brain Plasticity Lab,ESPCI Paris

ESPCI Paris
Laboratoire Plasticité du cerveau - UMR 8249
10 rue Vauquelin
75231 Paris Cedex 05
Tel : +33 1 40 79 47 56
Fax : +33 1 40 79 47 57
www.espci.fr

UNIWERSYTET JAGIELLOŃSKI
W KRAKOWIE

Dr hab. Anna Błasiak, prof. UJ
Zakład Neurofizjologii i Chronobiologii
Wydział Biologii
Uniwersytet Jagielloński

Kraków, 26.03.2025

Review of the doctoral dissertation by Konrad Danielewski

entitled

„Ensemble encoding of self- and other-related behaviors in the medial prefrontal cortex"

The doctoral dissertation „Ensemble encoding of self- and other-related behaviors in the medial prefrontal cortex", concerns the neuronal mechanisms underlying the social transfer of information. This is a significant and still insufficiently explored topic, as understanding how individuals interpret the behaviors and emotional states of others is essential for revealing how both humans and animals function in social environments, especially when confronted with potential threats. Notably, as the Author emphasizes, social interactions have the capacity to influence fear responses, potentially offering natural pathways to reduce fear through interpersonal engagement.

The dissertation begins with a list of the most commonly used abbreviations, which may not be a crucial part of the work, but would certainly benefit from being arranged in alphabetical order and slightly expanded in terms of content.

The Introduction is a very well-written part of the dissertation, in which the Author presents the key issues related to the social behaviors of rats in a clear and accessible manner, maintaining both relevance to the subject and consistency with the content of the subsequent chapters. Throughout the Introduction, the behavioral repertoire of the rat—the experimental animal used in the later studies—is defined and supported by well-chosen illustrations. The Author rightly emphasizes that social play and interactions during adolescence are essential for healthy social and neural development in these animals. Understandably, and in line with the aims of the study, the focus remains on the rat behavior, which is further justified by the limited number of studies on animals in this area. However, it would add further value to this section to briefly acknowledge—supported by relevant literature—that similar processes occur in humans and are equally critical for their social and neural development.

In the following paragraphs of the Introduction, the Author discusses the role of the medial prefrontal cortex, presenting it in a concise - perhaps even overly concise - manner. As a reviewer, I appreciate the overall brevity of the dissertation; however, it seems justified

to expand this section somewhat, particularly regarding studies focused on the anterior cingulate cortex (ACC). As the Author notes, the ACC plays a significant role in observational learning, the subjective experience of pain, and the anticipation of aversive events. Given the subject of the dissertation, it would be worthwhile to explore these aspects in more detail and support them with relevant literature.

The Author stated (already in the Abstract) that the aim of the study was to elucidate the role of the medial prefrontal cortex (mPFC) in social interactions, by analyzing neuronal activity in specific subregions of the mPFC during both the observation of emotionally charged behaviors exhibited by a conspecific and the expression of such behaviors by the experimental animal. Later, in "The aim of the dissertation" section, more specific research objectives are outlined. These include addressing whether the behaviors of self and others can be decoded based on mPFC activity, whether these behaviors are represented by the activity of individual cells or by population-level dynamics, and whether there is evidence of mirroring in the rat mPFC at either the single-cell or population level. While the specific objectives outlined in the dissertation were largely achieved, the broader aim presented in the Abstract remains unmet. The role of the medial prefrontal cortex (mPFC) in the studied behaviors was not directly tested. The dissertation describes the activity of specific mPFC subregions—at both the single-neuron and population levels—without experimentally verifying the functional involvement of the mPFC itself.

The Methods section presents the applied methodology in a generally clear and adequately detailed way. To conduct the research described in the doctoral dissertation, the Author employed some of the most advanced electrophysiological techniques available, within an exceptionally challenging experimental setting. Electrophysiological recordings were performed using Neuropixels probes in freely moving animals. Additionally, the analysis of electrophysiological signals was combined with video analysis. To the best of my knowledge, no Polish research group has yet published an experimental study using Neuropixels technology, which makes the presented results all the more noteworthy and deserving of recognition.

Given the applied protocol, one point that raises questions—and on which I would like to ask for the Author's comment—is the choice of an experimental paradigm in which the partners underwent a fear extinction protocol, but the experimental animals did not. Considering the key question posed in the dissertation—Does observing another's behavior recruit specialized cells or cell populations?—one might expect that a more appropriate design would involve a situation where the partner animal being observed by the experimental subject displays clear, emotionally salient behaviors, particularly those related to fear. This seems especially relevant given that the brain area under investigation, the medial prefrontal cortex (mPFC), plays a critical role in interpreting social contexts and regulating emotional responses.

The Author explains that the protocol was chosen because “the experimental design puts animals into a situation where meaningful information is shared between individuals due to similar experiences, i.e., the experimental animal has a reason to pay attention to its partner for information about the threat related to the CS.” However, at the same time, the partner has little reason to express emotional behaviors, having undergone extinction training, while the experimental animal has a strong reason to respond emotionally to the CS.

Given this, I would like to ask: does the Author believe that the experimental animal’s possible expectation of the unconditioned stimulus (having not undergone extinction) might influence mPFC activity, including the activity related to observing the partner?

The fact that the Author developed his own analysis tools, including training artificial neural networks and creating custom scripts for the analysis of electrophysiological data, is worth recognition. However, it is not clear how the electrophysiological signal was synchronized with the video signal.

The obtained results were subjected to in-depth analysis, including Principal Component Analysis (PCA) for testing feature collinearity. Additionally, a machine learning decoding approach was applied to analyze the behavior itself. This allowed for successful decoding of all examined behavioral categories and effective differentiation between events involving the experimental animal and those involving its social partner. The data were presented through various plots, trajectories, and heat maps, enabling a clear visualization of the findings. The analysis of neural activity revealed a lack of strong correlations and linear relationships between individual cells, indicating a high-dimensional nature of the data. However, the exact method used to assess correlations between features remains unclear—specifically, the type of correlation analysis applied, as well as how electrophysiological signals were quantitatively linked to behavioral variables such as velocity or acceleration; how electrophysiological signal was synchronized with video data. How movement acceleration was computed? Figure 25 is described in the text as showing the correlation between neural activity and acceleration; however, its legend suggests that it actually refers to velocity.

The discussion of the obtained results is conducted in a mature and thoughtful manner, demonstrating the Author's deep understanding of the topic and a strong ability to accurately interpret the findings. Importantly, the discussion provides clear answers to the specific questions posed earlier in the dissertation.

The Author clearly states that behaviors of self and others can be decoded based on mPFC activity, emphasizing the robustness of population-level decoding. Furthermore, based on the obtained results and in a well-justified manner, it is indicated that behaviors of self and others are primarily encoded at the population level, rather than by specialized individual neurons.

Finally, it is concluded that no sufficient evidence for mirroring in the rat mPFC was found—neither at the single-cell nor at the population level. The findings suggest that behavior can be understood without the need for single-cell mirroring mechanisms, and instead point to a broader integrative role of population-level dynamics within the mPFC. These are highly significant conclusions, pointing to new possibilities regarding the principles of encoding in the medial prefrontal cortex (mPFC).

In summary, Discussion is a very well-written section of the thesis. The only possible suggestion for improvement would be to consider linking the findings back to the Sherringtonian and Hopfieldian perspectives outlined in the introduction.

The bibliography consists of over 100 references, appropriately selected to match the content of the dissertation. However, there is a degree of inconsistency: the initial intent to organize the references alphabetically is not maintained in the latter part of the list. Additionally, some works cited in the main text are missing from the bibliography (for example, Paxinos & Watson, 2007).

In conclusion, I would like to indicate that Mr. Konrad Danielewski has presented a valuable doctoral dissertation, and the electrophysiological results contained therein contribute meaningfully to our understanding of the neuronal mechanisms underlying the encoding of behaviors of self and others in the medial prefrontal cortex (mPFC).

Minor comments, questions, or doubts raised in the course of the review do not affect the overall positive evaluation of the dissertation.

I hereby state that the reviewed doctoral dissertation entitled "Ensemble encoding of self- and other-related behaviors in the medial prefrontal cortex" meets the requirements specified in Article 187 of the Act of July 20, 2018 – Law on Higher Education and Science, and I recommend the admission of M.Sc. Konrad Danielewski by the Scientific Council of the Nencki Institute of Experimental Biology PAS to the next stages of the procedure for the award of the doctoral degree.

NTNU

The Kavli Institute for Systems Neuroscience
Centre for Neural Computation

Date: March 31, 2025

Assessment of PhD Thesis

Name of candidate:

Konrad Danielewski

Affiliation:

Laboratory of Emotions Neurobiology of the Nencki Institute of Experimental Biology, Polish Academy of Science

Supervisor: Ewelina Knapska

Auxiliary supervisor: Dr. Kacper Kondrakiewicz

Title of thesis:

Ensemble encoding of self- and other-related behaviors in the medial prefrontal cortex

Assessment Committee: *(name, title and workplace)*

External Reviewer:	Jonathan R. Whitlock, professor, Kavli Institute for Systems Neuroscience, NTNU, Trondheim, Norway

Evaluation of thesis

Evaluation of scientific quality of the dissertation and the quality of the data presentation in the thesis, evaluation of the organization of the dissertation.

The thesis is an independent and comprehensive piece of work investigating the neural representation of an animal's own behavior versus that of a conspecific, and the emotional benefit of social interactions between rats following feat conditioning. The work uses rodents as model species, which complements the majority of existing work, which is in non-human primates and humans. The scientific quality was high and the thesis adhered to high academic standards. The work contained in the thesis was experimental in nature, consisting of *in vivo* electrophysiological recording experiments in awake, behaving rats. The data were presented clearly and logically in the thesis, and the results were also presented clearly. The thesis followed a standard organization consisting of Abstract, Introduction, statement of aims, Methods, Results, Discussion, Conclusion, Bibliography and Publication record.

Evaluation of the scientific content

Address
Medical-Technical
Research Centre
NO-7489 Trondheim

Org.no. 974 767 880
E-mail:
whitlock@ntnu.no
http://www.cbm.ntnu.no

Location
Olav Kyrres gate 9
NO-7489 Trondheim

Phone
+ 47 45 16 43 90
Fax
+ 47 73 59 82 94

Date: March 31, 2025

The aims of the Ph.D. student's project:

The aims of the thesis were to address the questions of whether first- and/or third-person behaviour is encoded / decodable from neural activity in the prefrontal cortices (including the anterior cingulate, prelimbic, infralimbic, ventral orbital cortices); whether such features were represented by single cells, or at the neuronal population level; whether such representations depended on "mirror"-like tuning, such that neurons encoded the same actions whether they were executed in first person or merely observed.

This is an interesting topic on which to focus, since the cellular mechanisms underlying social cognition remain largely elusive, particularly since the volume of new results on mirror cells has declined in the primate literature, and the extent and reliability of mirror-like tuning in rodents is patchy. For example, previous work has provided evidence of mirror-like tuning in rats in relation to pain processing (e.g. Carrillo et al., *Current Biology*, 2019) and reaching / eating (Viaro et al., *Current Biology*, 2021), but both studies focussed on the tuning properties of single cells, so the question of mirroring at the population level is still open. The aims of the thesis are therefore well-motivated, especially considering the recording tools (e.g. Neuropixels) available in rodents.

In the following sections I give a summary of each section and include some points that would be good for discussion during the defence.

The employed research methodology

The study used adult male Wistar rats, housed in pairs and on a 12h light/dark cycle. Rats are perhaps the best choice of available rodent species, since they are both gregarious and are usually attuned to the environment and other animals in it. Habituation, handling, and the behavioral paradigm were clearly explained; and the fear-conditioning & extinction paradigms are widely used in the field as well. The surgical procedures, scoring of behavior, pose estimation, electrophysiology, spike sorting, dimensionality reduction and histological methods were clearly explained. It was also appreciated that, due to the nature of the fear learning and social conditions, that the window for collecting data was finite; it is indeed well-recognized that the approach is ambitious and difficult for obtaining in vivo recordings.

Questions from this section:

-Was the choice of Wistar rats deliberate, or based on availability? This is because Long Evans or other strains have pigmented eyes and therefore have superior vision.

-Why were the animals recorded from during their light cycle (i.e. when the animals would normally be sleeping)?

-A minor question has to do with which parts of the rats were tracked (using DeepLabCut)—it was not described which parts were tracked or why, though it was clear that the approach would give location, heading and other aspects of the animals' behaviour.

-I am personally curious as to how the 2-step surgery approach worked out in terms of cell yield for the NXP-1.0 implanted rats, and why it was not necessary for the NPX-2.0 implanted animals (i.e. if the multiple shanks on NPX 2.0's gave some kind of advantage).

The presented results

The results were generally clearly presented and were well-explained; consideration of the anatomical location of the recorded units (to the extent it was possible) was appreciated. The behavioural results (e.g. initial freezing of the experimental animals early in the test session; rearing in partner animals; Figure 7,

then Figure 9) were clearly presented, as the table summarizing how many units were recorded was useful. The spike rasters and heatmaps showing z-scored firing rates were clearly presented (Figures 12-13). The reported hit-rate for mirror-like neurons (3%) was quite similar to the hit rate in a prior study from my own; we found that such hit-rates could also be obtained using shuffled data; that could also prove useful in this work. The Venn diagrams summarizing the single-cell results were clear, but again that there was little overlap between neurons encoding performed and observed behaviors, which transitions the results into the population analyses.

The PCA results were again clearly motivated and explained well, and showed substantial differences in the neural state space for during the observed and performed behaviour. The decoding analyses (Confusion matrices) were perhaps the most crucial portion of the thesis, since they showed that neural activity during both performed and observed behaviors (regardless of how sparse) could be used to predict the occurrence of like (but not unlike) events during the recordings. Moreover, the interpretation that information is represented among the co-activity of the population is supported by the "subtraction" analysis, showing that decoding stays robust even when the strongest-coding neurons were removed.

Finally, the last three analyses (Figures 25-27) show that the correlation between kinematic features and pose-related features was weak, interpreted as meaning that spiking activity in PFC represented abstract states, rather than kinematics. This seems quite reasonable, but it was not entirely clear *which aspects* of the animals' behaviour that the analysis captured; the kinematic models of the animals could have been given some more explanation.

Some questions remaining from this section:
-A small point- to directly state in the results at which day the behavioural data were considered (e.g. in Figures 7-9)—though in the methods it was clear that this was on Day 5 in the paradigm.
-Was any analysis done to show drops in the experimental animals' freezing rates upon bouts of social interaction (like those shown in Figure 8)?
-What does the extinction rate look like for experimenter animals that don't have a conspecific on the other side of the chamber? What is the evidence that having the 2nd animal (partner) present impacts the experimenter animals?
-Would it be possible to see other examples of single cells encoding the behaviour of the partner animal? The examples shown had a bit sparse or delay spiking relative to the behaviour of the partner.
-Were the differences in PCA trajectories between performed and observed behaviors inevitable since the firing rates were quite different to begin with? Could this be done again with the same number of spikes used in each condition?
-How were neurons defined as being most "important" in the 2nd paragraph on page 49?
-Rearing, as a behaviour, was strongly decodable from neural activity, yet the kinematic features (which should include those for rearing) were not—how is that explained?

The discussion and conclusions

The discussion and conclusions were balanced, and considered both mPFC encoding both the animals' own and observed behaviors, as well as the observation that mPFC neurons appeared to preferentially encode more complex behaviors but not simple kinematics- reflecting the mPFC's role as an executive control centre, and one that is sensitive to the behavioral context. Other key aspects discussed included the primarily population-level coding of behavior in mPFC, which comports well with the primate literature, and the advantages and disadvantages of large-scale recording methods and unsupervised behavioural analyses

which are quickly becoming the norm in systems and behavioural neuroscience.

The choice of literature cited
The literature cited was appropriate and sufficiently broad to ground the thesis topic.

Information about detected mistakes/shortcomings, error and wrong or inaccurate wording
The writing in the thesis was clear are largely without mistakes, but I did catch a few typos along the way (e.g. 2nd to last paragraph on page 55, last sentence, should read "...as well AS by low confusion level...."). Theses were minor.

Assessment of whether the dissertation provides an original solution to a scientific problem
Yes, the core finding of the thesis, that the neural correlates for both performed and observed actions is strongest at the population level in the rat prefrontal cortex, is an original contribution (solution) to a scientific problem (i.e. that single-cell coding of such features is weaker).

Assessment of whether the doctoral dissertation demonstrates the PhD student's overall theoretical knowledge of the field and ability to conduct independent scientific work

The doctoral dissertation demonstrates that the PhD candidate has a good overall theoretical knowledge of their field, and that the student is able to independently conduct scientific work, including data collection, analysis, interpretation and writing.

Other comments
N/A

Overall evaluation
The overall evaluation of the doctoral dissertation is positive, and it is recommended that the doctoral candidate be admitted to the subsequent stages of the doctoral defense.

Suggested revisions

None that meet the criterion of "extraordinary circumstances", so no revisions are needed.

Trondheim, Norway, March 31st, 2025

Jonathan R. Whitlock, PhD
Professor
Kavli Institute for Systems Neuroscience
Norwegian University of Science and Technology (NTNU)
Olav Kyrres gate 9, 7030 Trondheim, Norway